(Meta)Systems as Constraints on Variation—A classification and natural history of metasystem transitions

Francis HEYLIGHEN*

PO, Free University of Brussels, Pleinlaan 2, B-1050 Brussels, Belgium

E-mail: fheyligh@vnet3.vub.ac.be

ABSTRACT. A new conceptual framework is proposed to situate and integrate the parallel theories of Turchin, Powers, Campbell and Simon. A system is defined as a constraint on variety. This entails a 2 × 2 × 2 classification scheme for "higher-order" systems, using the dimensions of constraint, (static) variety, and (dynamic) variation. The scheme distinguishes two classes of metasystems from supersystems and other types of emergent phenomena. Metasystems are defined as constrained variations of constrained variety. Control is characterized as a constraint exerted by a separate system. The emergence of hierarchical systems is motivated by evolutionary principles. The positive feedback between variety and constraint, which underlies the "branching growth of the penultimate level", leads to the interpretation of metasystem transitions as phases of accelerated change in a continuous evolutionary progression toward increasing variety. The most important MST's in the history of evolution are reinterpreted in this framework: mechanical motion, dissipative structuration, life, multicellular differentiation, sexuality, simple reflex, complex reflex, associating, thinking, metarationality and social interaction.

KEYWORDS: system, metasystem, supersystem, evolution, emergence, constraint, variety, control, hierarchy.

Introduction

It is an old, and widely accepted, idea that the systems around us can be analysed in a hierarchical way, characterized by different levels of complexity or organization (see e.g. Pattee, 1973; Mesarovic, Macko & Takahara, 1970). Since Darwin, scientists have also assumed that complexity is a product of evolution, and that systems characterized by a higher level of complexity have appeared following those with a lower level of complexity. However, few people until now have examined the connection between both ideas, and analysed the evolutionary development of hierarchical levels.

* Senior Research Assistant NFWO (Belgian National Fund for Scientific Research)

Herbert Simon (1962) is one of the early visionaries. He proposes to explain the hierarchical embeddedness of systems and subsystems, which he calls the “architecture of complexity”, by a process of random combination of elements, and the natural selection of those combinations that are stable. The selected combination forms a first level system, and can now again act as an element or building block, which is to undergo combinations with other elements, possibly resulting in a supersystem at a yet higher level of complexity.

Simon’s argument, however, explains only one type of hierarchical relation, which we might call *structural*, namely the one relating part to whole, or system to supersystem. Another type of hierarchy might be called organizational or *functional*: it describes the situation where a system at one level controls or directs a system at the level below. Whereas the structural hierarchy is typical for the physical world, describing the sequence from elementary particle to nucleus, atom, molecule, crystal, rock, planet, solar system, galaxy and supergalaxy, the functional hierarchy characterizes the world of life and mind, expressing the sequence from cell to plant, animal and human.

William Powers (1973; 1989) is one of the people who proposes a detailed model of such a functional hierarchy, inspired by evolutionary ideas. He starts from a very simple control scheme where a perception is compared with a goal (“reference value”), and the difference between them determines an action. The action is meant to compensate the deviation of the perceived value from the reference value, so that the perception would be brought as close as possible to the reference value. A hierarchy of control is easily constructed by assuming that the reference value of one level of control is set or controlled by the actions of the next higher level.

Another theorist who proposed an evolutionary account of the functional hierarchy is Donald Campbell (1974). He sees a control system basically as a *vicarious selector*, selecting the actions of the controlled system at the level below, in anticipation of the natural selection by the environment. It is this anticipatory weeding out of inadequate actions that makes the system more likely to survive in an uncertain environment. The functional hierarchy of control then becomes a “nested hierarchy of vicarious selectors”, where a higher level selector selects the selectors of the level below. This description is more general, but also less precise, than Powers’s scheme for the control of a reference value by a higher order reference value.

Though both theorists assume an evolutionary development based on blind-variation-and-natural-selection, which gives a clear advantage to systems characterized by higher levels of control, none of them proposes an explicit account of the process where a higher level of control emerges. The first one to have done that seems to be Valentin Turchin. He coined the word “metasystem transition” (MST) to describe the apparent evolutionary jump from one level to another one.

Turchin’s description (1977, this issue) of an MST contains both a structural aspect, like in Simon’s model, and a functional aspect like in Powers’s and Campbell’s models. The *structural* definition sees a metasystem S’ as an integration of a number of

subsystems S_i (which are normally replicas of some template S, like a single cell that is replicated in order to form a multicellular organism), to which an unspecified mechanism C is added that controls the behavior and replication of the S_i. This is similar to Simon's description, except for the additional control and replication mechanism. In Turchin's *functional* description, an MST takes place when the activity at the highest control level of some system S becomes itself controlled, forming a higher order system S': *control of S = S'*. This is similar to Powers's view, except that Powers's control scheme is more explicit (but because of that also more restrictive) on what is controlled and how control operates.

Apart from the explicit introduction of the MST as a fundamental process or "quantum" of evolution, describing evolutionary progress as a stairway of subsequent levels, Turchin's most important contribution is probably his "law of the branching growth of the penultimate level". This law might be seen as the beginning of a more detailed dynamics of MST's. It states that after the formation, through variation and selection, of a control system C, controlling a number of subsystems S_i, the S_i will tend to multiply and differentiate (instead of staying fixed, like in Simon's model for the formation of a supersystem). The reason is that only after the formation of a mechanism controlling the S_i it becomes useful to increase the variety of the S_i. Complementarily, the larger the variety of S_i to be controlled, the more important it is to develop and refine the control mechanism C. The development of C and the multiplication of the S_i are thus mutually reinforcing processes. The result is that an MST is characterized by a positive feedback where a small evolutionary change is strongly accelerated, after which it slows down again by the time a new level of equilibrium is reached.

Turchin's alternation between structural and functional definitions makes the understanding of an MST to some extent ambiguous. Depending on the concrete example, Turchin uses either a structural description (like in the emergence of multicellular organisms), or a functional one (like in the emergence of learning or associating as the control of reflexes), or some mixture of both. It appears as though Turchin has consciously kept the definition of an MST vague, so that a maximum of phenomena could be described by it. Most recently (Turchin, this issue), he has even interpreted the formation of a molecule by integration of atoms as an MST, thus leaving the domain of living and technological systems to which the phenomenon of "control" is traditionally restricted.

Though there may be an advantage in having a very broad definition of MST, that would even encompass Simon's "structural" scheme for the development of a supersystem, it would be clearly useful to make at least a classification of different types of MST's, where each type would be defined in a much more restrictive way. This would allow the elimination of possible ambiguities and confusions between the different approaches and concepts mentioned, such as supersystem vs. metasystem, control vs. selector, or control by integration vs. control by setting of reference values. A more precise description of different MST-types might also make it easier to develop a detailed

dynamics of MST's, which could function as a tool for predicting or directing concrete processes.

The present paper will attempt to develop such a classification, by going back to the most primitive level, using a new definition of the concept "system" as a basis for an alternative definition of "metasystem" and "supersystem". Though the resulting characterization of MST will be different from Turchin's, it will be shown that most of their practical instances overlap, so that the difference is more one of form than one of content.

System as constrained variety

Though there about as many definitions of "system" as there are people who have thought about the issue, it is possible to distinguish two broad classes of interpretations. A first, constructivist or subjectivist, position sees a system as anything which is distinguished by some observer as a system. The more traditional, "objectivist" view sees as system as a set of parts that are somehow interrelated so that they form a whole (see Joslyn, this issue, for more details).

What both definitions have in common is that a system must at least have some structure or distinction, allowing it to be separated from its background or environment. Starting from that feature of distinction, I would like to propose a definition that is more precise than the subjectivist one, yet more general than the objectivist one. What I believe to be lacking in the subjectivist one, is that there is not only a distinction between system and non-system, but also a number of distinctions internal to the system itself. It is this internal structure that makes us recognize something as a system, rather than as a simple appearance or sensation.

In the objectivist definition, the internal distinctions are produced by the parts, viewed as spatially separate, static phenomena. But this is too restrictive. In physics, an elementary particle, for example an electron, is called a physical "system", even though it does not have any distinguishable parts. Yet, it does have distinct properties or states: for example, it can have different positions, energies, or momenta. In spite of its various appearances, the electron is still considered to have a stable identity, since the different appearances are connected by a continuous evolution of the electron's state vector, governed by physical laws, as expressed by the Schrödinger equation.

An example from cybernetics might be the simplest type of thermostat, which is defined by its two states, "temperature too low" and "temperature high enough", and its rules for selecting actions that would maintain or bring the state back to its preferred value. In this "systemic" characterization the internal parts or structure of the thermostat are ignored in favor of its functional organization.

What these two elementary systems have in common with more complicated systems consisting of many parts is that they can undergo a variety of appearances, while

maintaining an invariant identity. The identity corresponds to the “external” distinction between system and environment; the variety of states or appearances corresponds to a set of internal distinctions between those states. The maintenance of the identity implies that the variety of aspects or appearances must somehow be restricted: if a system could take on any appearance at any time, it would be impossible to distinguish it from another system, since the latter might take on the same appearance. (In the case of electrons, this self-evident principle even takes the form of a physical law, the Pauli exclusion principle, which states that no two electrons can be in the same state at the same instant.)

This necessary connection between “external” identity and “internal” distinctions leads me to redefine a system as a *constraint on variety,* where the constraint determines the invariant identity in spite of the variety of appearances. The variety describes the fact that a system necessarily has some internal differentiation, proposing distinct parts, aspects or states. The constraint expresses the condition that this variety must somehow be restricted, so that the number of possible appearances at a given moment is limited, allowing us to distinguish appearances within those limits as belonging to the given system, from appearances outside those limits as belonging to something different.

A related concept of system is used in mathematical systems theory (see e.g. Mesarovic & Takahara, 1975). Here a system is defined as a subset of the set of all possible input-output connections (Cartesian product of set of possible inputs with set of possible outputs). The total set of connections might be interpreted as a maximal possible variety, the subset defining a system as a limited variety, to which actually occurring input-output transitions are constrained.

However, in my view, a formalization of the notion of constraint, as determining a system’s identity, would require a more specific construct than that of a mere subset. Not every subset of a variety of possible configurations should define a system. A stronger requirement is that it should somehow be impossible for the system or process to leave this subset, as if there were a boundary blocking further movement. A better concept to express this intuition is that of a “closed” set, which entails a clear distinction between inside and outside of the set, and which can be mathematically defined (see Heylighen, 1990). Closure of a set also expresses a feature of invariance or stability of the system’s identity. We will not go into further details about this idea, but note that the term “constraint” in its present sense is more or less equivalent with the term “closure” as I have used it in earlier papers (Heylighen, 1989, 1990, 1991a,b).

Static and dynamic variety

When we speak about the variety of appearances of a system we should distinguish two basic types of variety: *static* and *dynamic.* The static variety corresponds to the segmentation in parts, which can be distinguished independently of any change or evolution in the system’s configuration. The variety of parts leads to a variety of

appearances when the system is observed from different angles or points of view, because different parts will come into focus.

An example of a system with static variety would be a crystal, where distinct molecules are arranged in different positions, but where both the types of molecule and the positions are constrained in such a way that we can unambiguously determine whether a certain molecule belongs to the crystal or not. More abstract examples can be found in formal systems, such as the set of natural numbers or the periodic table of elements. Both contain separate elements, which follow certain rules, so that elements of the system can be distinguished from non-elements, which do not obey the rules.

Dynamic variety appears when a system undergoes subsequent states or configurations during its temporal evolution, and this independently of the fact whether it has a static variety or not. For a given point of view, there is now still a variety of possible appearances through which the system might pass during its evolution, and this is what we will call dynamic variety. The process of "visiting" those potential appearances will be called *variation.*

The electron is an example of a system with a purely dynamic variety, that is to say without static variety of parts. The periodical system of elements, or a crystal fixed in the crust of the earth so that it cannot move or undergo state transitions, are examples with static variety but no dynamic variety.

Most practical systems have both static and dynamic varieties, though, and that often leads to confusion. Static and dynamic variety are linked in those (most usual) cases where the parts of a system can "move", or, more generally, undergo changes, relative to each other. This is the case when the parts have a dynamical variety of their own. In that case, the overall dynamic variety of a system is equal to the product of the dynamic varieties of its parts.

A classification of supersystems and metasystems

When defining higher level systems, the intuition we try to express is that of a system which is "above", "about" or "containing" some other system(s). Perhaps the simplest concept to start with is that of a supersystem, which might be viewed as a "system of systems". With the above definition of a system, this becomes:

a supersystem is a "constrained variety of systems", i.e. a "constrained variety of constrained varieties". (1)

Since the supersystem has its subsystems as parts, we will interpret the variety of the supersystem as static variety (which may lead to dynamic variety if the subsystems are dynamic). With a metasystem, on the other hand, our intuition would see it not as a mere static collection of systems, but rather as a system that somehow controls, directs or

manipulates the systems at the level below. One way to express that intuition is by replacing the static *variety* of definition (1) by dynamic *variation*. We then get:

a metasystem is a "constrained variation of (a) system(s)", i.e. a "constrained variation of constrained variet(y)(ies)". (2)

What does this mean concretely? "Constrained variation" might be paraphrased as "constraint on variation". The last part of definition (2) then becomes: ... variation of constraints on varieties. So what is varied in the metasystem are the constraints that define the systems of the level below (which we will call "object systems"). The metasystem manipulates the identity of the object systems themselves.

This comes close to the traditional use of the prefix *meta-*, as in metamathematics or metalanguage. A metalanguage is not a mere collection of languages (that would be a "superlanguage") but a language that can be used to make assertions about another, "object" language, thus defining the form and content of the object language. Similarly, a metamathematical expression is an expression that allows the manipulation (generation, deduction, testing) of mathematical expressions at the object level. Such an expression (e.g. a deduction rule, like the modus ponendo ponens, expressed as a formula) can be interpreted as representing a constrained variation of object expressions: it allows the generation of a limited number of new expressions, starting from the given expressions.

Though supersystem and metasystem are both "systems on systems", they are fundamentally distinct by the fact that in the former the variety is static, in the latter it is dynamic. Like we noticed before, in the general case there is both static and dynamic variety. In such a case the resulting second level system might still be called "metasystem", since we tend to assume that the dynamic aspect is more important than the static one. This corresponds to Turchin's structural definition of a metasystem, where there is both an integration of subsystems (constrained static variety, similar to Simon's supersystems), and a control manipulating them (constrained variation, similar to Powers's model).

The issue may be clarified by introducing the concept of "scale" of an MST (Turchin, this issue). The scale n is the number of systems that are integrated. It corresponds to the static variety of the higher level system. In order to have a supersystem, the scale is necessarily larger than one: $n > 1$, otherwise there would be no variety in definition (1). In the case of a metasystem, though, definition (2) allows both $n = 1$, and $n > 1$. In the former case the variety is purely dynamic. An example would be a thermostat (first level control system) that is controlled at the metalevel by a timer, that sets the reference temperature depending on the period of the day. The variation of the temperature is constrained by the thermostat's setting, but that setting itself undergoes variation during the day, constrained by the timer mechanism.

Incomplete metasystems

In order to avoid confusion it is useful to consider incomplete cases of definition (2), which may look like super- or metasystems, but which are not, since one or more of the essential second order components (constraint, (static) variety, (dynamic) variation) is lacking. The different possible combinations of constraints and varieties can be summarized by a 2 × 2 × 2 scheme (see Table 1.), which classifies all 8 possible combinations of:

{constraint, no constraint} on ({dynamic variety, no dynamic variety} and {static variety, no static variety}) of constrained variet(y)(ies).

Dynamic Variety	**Static Variety**	**Constraint** *yes*	*no*
yes (flexible)	*yes (scale>1)*	Metasystem or... (see text)	Aggregate of relaxed constraints
	no (scale = 1)	Scale-1 Metasystem	Relaxed constraint
no (rigid)	*yes (scale>1)*	Supersystem	Aggregate
	no (scale = 1)	Additional constraint	null case

Table 1. A 2 × 2 × 2 classification scheme for metasystems and related higher order systems.

First, a *variation of constraint on variation* (Row 2, Column 2). This would mean that the constraint defining the system at the object level is arbitrarily or unrestrictedly changing. The result is that the variation at the object level can take on different values depending on the arbitrary constraint that is there at that moment. The net effect is indistinguishable from an object variation without constraint or with diminished constraint. The metalevel variation merely annihilates or *relaxes* the object level constraint, without adding any higher form of organization. A physical example of the emergence of such a phenomenon is the evaporation of a liquid: in the liquid the molecules can move, but are constrained to a given volume. In the gas formed after evaporation, that constraint has disappeared, and molecules can move freely. This is equivalent to saying that the volume wherein they move (original constraint) has become variable.

Second incomplete definition, *constraint on constraint on variation* (Row 4, Column 1). This is merely an additional constraint, restricting the variation at the object level even further. To continue with our example of the liquid, an added constraint might appear through freezing: in the frozen state not only the volume, but also the positions of the molecules have become fixed.

Third, *variety of constraints on variety* (Row 3, Column 2). This is not a supersystem, but merely a variety or “aggregate” of independent systems. For example, sand is an aggregate (unconstrained variety) of grains of sand. If the unconstrained variety is accompanied by unconstrained variation, we have an aggregate of systems with relaxed constraints (Row 1, Column 2). An example might be a mixture of gases.

Finally, there is the null case where neither constraint nor variety are imposed on the existing systems (Row 4, Column 2). In that case, no new or higher order system can be distinguished.

For the sake of completeness, we must remark that in this scheme (and in Table 1.), we have only considered the presence or absence of constraint on variety in general. In the case where there is both static and dynamic variety, there might be a constraint on either or both types of variety. In the case where there is constraint on variation but not on variety, we get an aggregate of metasystems with scale 1. Constraint on variety, but not on variation, leads to a supersystem consisting of systems with relaxed constraints. It is only when both types of variety are constrained in a coherent way that we have a true integrated metasystem.

Control as external constraint

Our presentation until now has defined MST’s by means of the general term “constraint”. Turchin’s original definition, however, as well as Powers’s more specific hierarchy, use the concept of “control”. Campbell, finally, speaks about “(vicarious) selectors”. In order to avoid confusion, we must clarify how the different terms are related. (A related analysis of the concepts of constraint and control is proposed by Joslyn, this issue).

Though control is certainly a form of constraint, the basic difference seems to be that control requires a *controller*, that is to say a system physically or structurally separate from the system being controlled. In the concept of constraint, it is not necessary (though it is possible) to situate the constraint in a separate system, the “constrainer”. A constraint may be inherent in the system being constrained: the system may simply be the result of a natural selection, where unstable configurations have been eliminated so that only a restricted or constrained set of configurations remains. The concept of an “attractor”, as a region in state space that a system can enter but not leave, exemplifies such a spontaneously arising constraint. For example, in a crystal the molecules are constrained to positions on a rigid grid, but there is no force outside the molecules themselves that is responsible for keeping them there.

On the other hand, the fact that control resides in a separate system, implies that there must be channels of interaction between controller and controlled. These channels will not be perfect. That means that information will not be transmitted completely or instantaneously. It implies that a variation of the controlled system cannot be exactly constrained: there will always be a delay between the start of the variation and the reaction

of the controller constraining that variation, and in the meantime the system may have left the domain of configurations to which it was supposed to be constrained. Moreover, any noise or loss of information over the channel implies that the reaction will not be perfectly adequate to counteract variations leaving the domain of constraint.

In spite of these inherent limitations, control can be very effective. This is made possible by the negative feedback relation between controller and controlled. As demonstrated eloquently by Powers (1973; 1989), it suffices that the reaction of the controller to any deviation of the controlled system from its domain of constraint be opposite to and larger than the deviation itself for the system to be very stable. However, it does not seem necessary to postulate, as Powers does, that the domain of constraint must be limited to just one "reference" configuration, or that deviations can be exactly measured by one dimensional quantities, or even that for every given deviation, the controller can unambiguously determine the adequate counteraction.

Campbell's approach, though less detailed and precise than Powers's, provides a more general picture of the mechanism of control. In Campbell's terminology, the controller becomes a (vicarious) *selector*. The selector merely tries to eliminate variations leaving the domain of constraint, but it can fail. It functions as a heuristic device, for finding an appropriate counteraction by vicariously testing different possibilities. There is no deterministic one-to-one or many-to-one mapping from perturbation to selected counteraction, like in classical (and Powersian) control systems, as the variation undergoing selection is essentially blind. The selector is seen as a kind of template in which variations must fit, but the mechanism of selection is not further specified.

For Powers, there is control rather than constraint (e.g. a stable equilibrium or attractor) when a system is maintained in a configuration that is not stable on its own. This is a special case of our definition, since it implies that there must be a force or constraint *external* to the system that keeps it away from its internal equilibrium state. But one might wonder why a system should be maintained by a controller in a situation that is inherently unstable. The reason is that there may not be a globally stable configuration, because the environment is too complex and variable to leave equilibria undisturbed.

The advantage of an external controller/selector is that, being separate from the controlled system, it can undergo variation without therefore implying a variation of the controlled system. That means that the controller can somehow adapt, learn or—as Powers call it—reorganize, i.e. improve its functioning, without interfering directly with the system it controls. This is not possible for an internal constraint, since such a constraint derives from the dynamics of the system itself: any change of the constraint, apart from a discontinuous jump to a new attractor, implies a change of the system's dynamics. With an external constraint, embodied in a separate selector or controller (e.g. the DNA as controller of cell metabolism, or the brain as controller of the body), on the other hand, small, almost continuous changes of the controller (e.g. mutations, or changes in neural connection strength) are possible without a major disruption of the controlled system. That makes it possible for the controller to quietly explore a large

variety of configurations, and retain those that make the complete system (controller + controlled) more stable.The adaptation of the controller may become constrained or controlled itself, resulting in an metasystem.

Another advantage of the separation between controller and controlled is that the control system can be shielded from environmental perturbations by the controlled system. For example, DNA is protected by the cell wall, and kept in shape by different mechanisms of self-repair using enzymes of the cell. The brain is encased in a hard skull, maintained on a constant temperature, and supplied with a constant influx of oxygen and nutrients by the body. This guarantees that single variations of the controller will be relatively small, so that the probability of major disruptions is minimized. The controller will typically vary much more slowly than the system it controls. The net effect is that the controller or vicarious selector becomes a kind of repository of information or knowledge, on how to react adequately to perturbations, which is stable in the short term, but capable to slowly adapt, and become better, in the long term.

The separation finally implies characteristic of controllers, which is captured by Campbell's (1974) term of *vicarious* selectors. The selections made by the external constraint are in a way "vicars", substitutes or representatives, of events of long-term, natural selection. Both types of selection eliminate "unfit" actions. The difference is that a vicarious selector can eliminate an inadequate action *before* the system is destroyed, that is to say it *anticipates* natural selection. An inherent constraint, on the other hand, functions "in real time": when an action leaves the domain of constraint of the system, the system by definition ceases to exist. No anticipation is possible. Only by separating constrainer and constrained is it possible to create a dynamics that moves faster than, and thus allows to anticipate, the dynamics of the constrained system. The "vicariousness" or anticipation is what makes a controller into a model (Turchin, this issue; Conant & Ashby, 1970) of the variation processes it controls.

In conclusion, separating a constraint from the system it constrains, and embodying it in a distinct controller or selector, makes the variation of constrainer and constrained to some degree independent. This has the following advantages: the controller can anticipate processes involving the controlled; the controller can evolve without disrupting the controlled; the controlled can function as a buffer or shield protecting the controller. The disadvantage is that noise and delays will weaken the communication, but this is more than compensated for by anticipation.

It must be noted that we interpret the separation between controller and controlled as an actual, material, spatial or structural, separation. It is always possible to *functionally* separate or distinguish two properties of a system, and call the one, e.g. the dynamical constraint, the "controller", and the other one, e.g. the particles being constrained, the "controlled". However, this functional separation does not entail the possibility of independent variation, which we consider here essential to distinguish control from constraint. It is this independence which makes it much more likely that the variation of the controller would hit upon a constraint itself, resulting in an MST.

Dynamics of variety and constraint

Emergence

After proposing a classification of static structures we need to explore the dynamics of systems and metasystems: what are the possible or probable transitions from one form of constrained variety to another one? I have defined an *emergence* as any process whereby the variety and/or constraint of a system change (Heylighen, 1991b). Given the above definition of system, such a process will necessarily change the identity of the system itself. It might therefore also be called a system transition. This is a qualitative change, where a new organization or system appears, with properties (potential appearances) that did not exist in the old system. The more usual ("quantitative") dynamical evolution of a system, on the other hand, is merely a transition within the constrained variety of possible appearances, where neither constraint nor variety undergo any change.

Physical phase transitions, like crystallization, freezing, melting, evaporation and condensation, are elementary examples of emergence in the above sense. Heating solid matter tends to destroy constraints, and simultaneously, increase variation, resulting first in a liquid, then in a gas, and finally in a plasma. Cooling down matter, on the other hand, decreases variation, allowing constraints to be reinstated. These transitions lead to configurations as found in the second column and last row of Table 1.

Physical or chemical reactions, where new particles or molecules are formed by recombination of parts of old ones, are more complicated forms of emergence, where constraint or variation do not monotonically increase or decrease. Rather, one form of constraint tends to be replaced by another one, which is not necessarily stronger or weaker, but which allows different types of variation. The resulting configurations cannot be situated in Table 1.

In order to get one of the configurations in the first 3 rows of column 1, we need a more specific type of emergence or transition. To explain how such transitions can happen, we must go back to some fundamental principles of evolution (Heylighen, 1992a).

Evolutionary principles

Perhaps the most basic assumption of the present evolutionary approach is the idea of variation and selective retention (cf. Campbell, 1974): configurations (systems or more primitive phenomena) continuously undergo changes, until a configuration is reached that is stable. This means that unstable configurations are eliminated and replaced by different (stable or unstable) configurations. Eventually only the stable configurations are retained. This spontaneous separation between stable and unstable systems may be called natural selection.

This approach assumes that there is already a variety and variation of configurations available. Constraint, on the other hand, arises automatically by selective retention. A stable configuration is characterized by the fact that its elements or parts do not undergo certain variations that would destroy the configuration. In other words, its remaining variation is constrained (the corresponding decrease of dynamic variety can be understood from the "principle of asymmetric transitions", Heylighen, 1992a). A stable configuration can, hence, be interpreted as a system, in our present definition. The transition from undifferentiated phenomena to systems is therefore a fundamental evolutionary process, which does not need much explanation.

Supersystem transitions

The transition to a meta- or supersystem, on the other hand, is more complicated, since it already necessitates the existence and availability of stable systems. The emergence of a supersystem is relatively straightforward. Assuming that we have a static variety of systems (stable configurations), the possible relations of these systems with respect to each other will determine a dynamic variety of possible configurations. The variation, moving among these configurations, will eventually hit on a configuration (or set of configurations) that is stable, and this will determine a constraint on the static variety of systems that fulfil the stability requirement. No more systems will be able to join or leave the configuration without destroying its stability. In that way a stable supersystem is created.

This is the process originally described by Simon (1962). He argued that the typical number of systems integrated in such a way would be rather small. Indeed, the larger the number of systems to be combined, the smaller the probability that they would hit a combination that is stable enough to maintain as a new system. However, Simon did not take into account the possibility that a small assembly of systems might, instead of immediately "closing its boundaries", make it easier for other systems to join the assembly. This happens for example during crystallization, when a small assembly of molecules with a stable geometric configuration acts as a template making it easier for other molecules to join the configuration. Such configurations are characterized by autocatalytic growth (Heylighen, 1992a), i.e. a positive feedback process that augments the static variety, with the only constraint that the added systems belong to the same general type that "fits" the template.

Metasystem transitions

The selective retention of a stable configuration depends essentially on two factors: the configuration must be intrinsically stable (internal selection), and able to resist changes in its environment (external selection) (cf. Heylighen, 1991b). It is the latter factor that influences the amount of remaining (constrained) variation of a system. Following the principle of selective variety (Heylighen, 1992a), we may state that the higher the internal

(dynamic) variety of a system, the larger the number of environmental situations in which it can maintain. Indeed, different system configurations will be stable in different external situations. In general, if the number of situations increases, also the variety of system configurations will have to increase for the system to keep the same probability of retention. Even for an environment with a given variety of situations, systems with a large internal variety will be more likely to survive, since the external variety of situations will normally always be larger than the internal variety of configurations. Thus evolution will tend to increase the internal variety.

This spontaneous growth of variety is reinforced through a positive feedback produced by the interaction between different evolving systems. As every systems increases its internal variety in order to better match the variety of its environment, it will thereby increase the variety of the other systems' environment (to which it belongs), and thus force them to increase their own internal variety even more strongly in order to keep with the higher external variety.

Internal variety cannot increase without limit, though. The larger the number of internal configurations, the more variation or trial-and-error the system will have to undergo before it finds the adequate configuration (i.e. the one that is stable for the given situation). If that exploration takes too long the system may already be destroyed before it finds the appropriate configuration. The variation process can be shortened by constraining it in such a way that unlikely configurations are not explored. (this might be formulated as a complementary "principle of requisite constraint or requisite knowledge", Heylighen, 1992a). But that seems to imply that the overall variety of the system would be diminished, and that goes against the principle stated above, which confers a selective advantage to increased variety.

How can we reconcile these two apparently contradictory requirements, one for increasing variety and one for adding constraint? In the simplest case the system will reach a trade-off or compromise, sacrificing some of the benefits of variety for some benefits of constraint. However, it is possible to increase *both* variety and constraint by shifting to the metalevel: constrained variation of the constraint that defines the system. The resulting metasystem will have a much larger variety of possible configurations, since the original variety of configurations at the object level must now be multiplied by the variety of constraints at the metalevel. Still, the variation process will be constrained at both the metalevel and the object level, allowing a relatively quick selection of the most adequate configuration.

We might conceive the emergence of such a metasystem in the following way. Assume that a system has insufficient internal variety to maintain in its given environment, and that it can only reach the adequate configurations by changing its defining constraint. If the constraint changes in a relatively slow, continuous way, we might say that the system survives the changes, even though its identity has changed in the process. But the variation of the constraint itself undergoes selection, and that means

that most likely an eventual stable configuration of second-order variations will be reached, defining a second-order constraint. This defines a new metasystem.

When this process is combined with the process of a supersystem transition, as described above, we will find that several systems undergoing second order variation collectively develop an overall constraint on their mutual variations. Here we come back to Turchin's original structural characterization of a metasystem transition, where a number of subsystems are integrated, under the control of a higher order mechanism of constrained variation. The advantage of integration of several systems is that systems which were originally similar (for example because they are replicas of each other, as in Turchin's original definition, or because they arose through similar circumstances) can now be varied independently. This creates a differentiation between the subsystems. An example is a multicellular organism where all cells were initially identical, but became differentiated in different organs and tissues. The metasystem may impose different (first order) constraints on its subsystems, yet maintain an overall (second order) constraint that co-ordinates their activities or configurations, so that the overall configuration is adapted to the situation. In that way the overall constrained variety is much larger than either that of a sequential metasystem without static variety of subsystems, or that of a supersystem without dynamic variation of constraints.

Branching growth of the penultimate level

Turchin's law of the branching growth of the penultimate level can be easily added to this framework. Once the metasystem (second order constraint) is established, because of the principle of selective variety, the variety of the system will tend to increase, until the limit is reached when the "principle of requisite constraint" comes to the foreground, i.e. when the existing second order constraint becomes insufficient to guarantee that the search for a stable configuration will not become too long. That variety can increase by increasing the number of subsystems ("branching growth"), or by making the first order constraints more variable (case not discussed by Turchin). Simultaneously, the increase of variety will tend to strengthen the requirement for an adequate constraint, and thus stimulate the development of the second order constraint.

Thus, development of variety and of constraint will tend to reinforce each other, up to the point where further increases of variety will be more difficult to adequately constrain at the given metalevel. At that moment, the process of further complexification of the system will slow down, and the MST may be said to have come to a halt. Evolution, however, will continue at a slower pace, until the moment where, possibly because of a change in the environment, a new MST is triggered. When variety is plotted in function of time, the resulting graph will be in general monotonously increasing, but the increase will be much faster during MST's (see fig. 1). Thus, although evolution could be seen as a continuous increase of total variety or complexity, a more coarse-

grained picture of the same growth curve would look more like a step function, with MST's as the discrete jumps from one level to a next one.

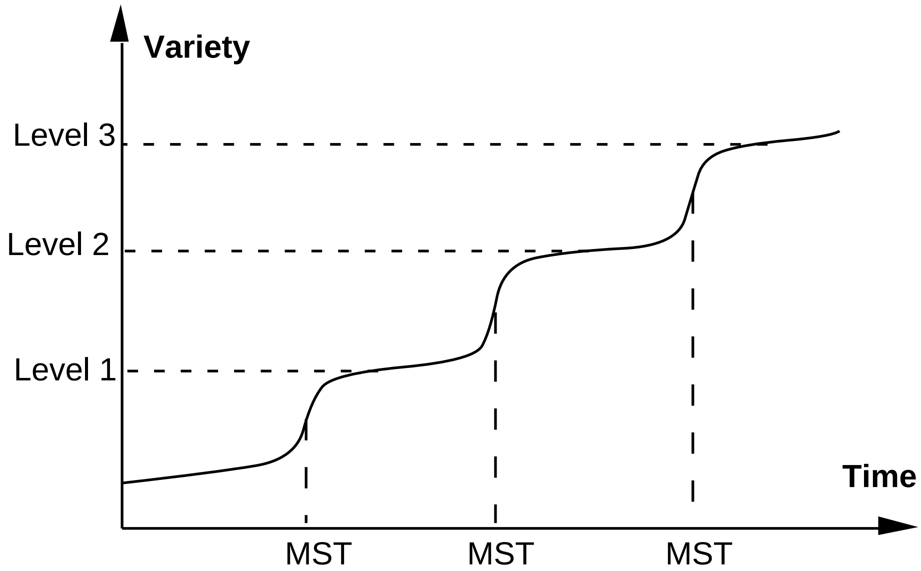

Fig. 1: a cascade of subsequent metasystem transitions, mapped as an increase of variety during temporal evolution.

The main sequence of MST's

We will now discuss some of the most important MST's in the evolution of the universe as we know it. We must first remark that there are many possible MST's, big ones as well as small ones, and that MST's may affect different part or aspects of constraints characterizing different systems. In particular this implies that it is possible to have parallel MST's, where the same system is incorporated in two different metasystems, each characterized by the variation of a different part of the constraint defining the original object system. However, such parallel metasystems hierarchies are rare, as it is not very likely that the variation of the same system would satisfy two independent selection criteria. The argument is similar to the one that led Simon (1962) to assume that a subsystem would belong to only one supersystem, rather than to several overlapping supersystems (cf. Heylighen, 1992a).

We will therefore examine one basic sequence of MST's in depth, and then quickly sketch some parallel sequences, that affect less important aspects of the system. The sequence will be similar to the one originally proposed by Turchin (1977), except that it will start at an earlier, prebiotic level.

Mechanical systems

The level we will start with will be that of mechanical or physical systems, such as particles, atoms and molecules. Such a system is defined as a constrained variation of

some elementary configurations, states or appearances. These are characterized by elementary properties or distinctions, such as position, momentum, spin, mass, etc.

Though mechanical systems are arranged in a hierarchy of their own, e.g. quark, nucleon, atom, molecule, crystal, ..., that hierarchy is of the subsystem-supersystem type, and not of the object system-metasystem type. Turchin (this issue) recently did analyse a molecule as a metasystem controlling its component atoms. But in the present definition, the relation is one of constraint rather than control, since the force binding the atoms together would not be seen as a controller external to the systems being controlled. The force field (represented in Turchin's model by the electron shell shared by the atoms) is part of the system "atom". So we would view the molecule as a constraint on its component atoms. The atoms themselves form an obvious static variety, so the requirement to have a supersystem is fulfilled. But in order to have a metasystem we should also conclude that the constraint defining an atom as a system is undergoing constrained variation. And that requires an analysis of elementary systems.

Mechanical systems (atoms, particles, molecules, ...) are characterized by a dynamic variety of possible states together with a constraint on transitions between those states that can be expressed as a set of causal laws (conservation laws, dynamical equations, ...). The constraint determines a trajectory of the system through its state space, depending on the state of the environment (fields, forces, or other systems present). As long as the state of the environment is not known, the trajectory will be undetermined, but still minimally constrained by the general laws (e.g. the trajectory must be continuous). The only way to vary the constraint is to vary the environment, thus influencing the trajectory.

Since the formation of a molecule by the chemical bonding of atoms restricts the variation of the atom's environment, it cannot be interpreted as a variation of the atom's dynamical constraint, and hence not as an MST in our present definition (though it might be one in Turchin's (this issue) definition). A similar reasoning applies to all forms of physical or chemical bound states, like the bonding of quarks in a nucleon, of electrons in an atom, of molecules in a crystal, or of planets in a solar system. All these systems are supersystems, not metasystems, with respect to their elements.

Dissipative systems

In order to build a real metasystem out of mechanical elements, we need to continually vary the "undisturbed" trajectory of the system, but such in a constrained way. The prototype of such a system seems to be what Prigogine (Nicolis & Prigogine, 1977) has called a "dissipative structure". In a dissipative structure there is a continuous, stochastic variation of the trajectories of the elements (usually molecules). This ongoing process requires an influx of energy, which is dissipated in the form of entropy. Yet the process obeys a higher order constraint, and the trajectories which seem random at the microscopic scale, appear to follow regular patterns at the macroscopic scale. Typical examples are the Bénard pattern of rolls or cells appearing in a liquid heated from below,

and oscillating chemical reactions, such as the Brusselator (Nicolis & Prigogine, 1977). Yet we do not need to consider such unusual phenomena to get a feel of what this first level of metasystems consists of. Everyday phenomena like rivers, flames, whirlwinds, vortices, ... are dissipative systems, in which a multitude of interacting elements vary each others trajectories or evolutions in such a way that a macroscopic constraint emerges, which cannot be reduced to the constraints on each individual trajectory.

What is typical of a dissipative system is that it can only survive in a specific environment where a constant flow of energy and new material through the system is maintained. A flame will stop to exist as soon as all its fuel is burned. A river can only maintain because the heat of the sun makes the water of the oceans evaporate, after which it condenses, producing rain that feeds the river's sources. In general, however, the amount of available energy and material will be variable. Most dissipative structures may not even survive small variations in their input and output (non-equilibrium boundary conditions). That is why a dissipative system in a variable environment will need an additional control mechanism to sustain its survival.

The origin of life

If different environmental circumstances sustain different dissipative structures, the system, in order to survive, must be able to vary the constraint, determining the structure. But in order to still have a recognizable system, the change of structure or constraint must not be random, but itself constrained. So we come to a metasystem of a second order, constraining the variation of the first order metasystem.

Such a metasystem is characteristic for all living systems. All organisms are continuously undergoing dissipative chemical reactions, producing energy, structure proteins, enzymes, and many other products, in a regular ordered way. Yet most of the parameters influencing these dissipative cycles, such as outside temperature, amount and type of molecules available, are changing constantly. In order to keep the metabolism going, the speed and type of reactions must be continually adapted to these changing circumstances. This is done through a selective release of enzymes, which catalyse specific reactions. The release of enzymes in function of the physico-chemical parameters is controlled by the DNA, which performs the role of selector or external constraint. Depending on the available proteins in the cell certain parts ("genes") of the DNA are either activated or inhibited to produce the requisite enzymes.

Definitions of life, entailing specific accounts of how life originated, can usually be classified in two categories: 1) *metabolic*, emphasizing the sustained cycle of reactions; 2) *replicative*, focusing on the capacity of DNA to reproduce itself. The present view would rather start from metabolism, but acknowledge the fundamental role of DNA as a stable, separate selector of metabolic processes. The replicative function is not essential to a definition of life as a second-order metasystem (you hardly would deny your grand-mother the property of being alife because she has passed the age of reproduction), but it

is more than an accidental property. Replication is not limited to living organisms: crystals and other physico-chemical structures can be said to be capable of self-reproduction. It suffices that some already formed stable configuration would function as a template, selector or catalyzer for the formation of similar (like in crystals) or complementary (like in single DNA strings) configurations out of freely available components.

But replication is a dissipative process, requiring a continuous input of new material. Therefore it is in the interest of replicators to take part in a dissipative system supplying the needed resources (cf. Powers's scenario for the origin of life, this issue). That is what happens in living cells, where the nucleotides and other components necessary to build a DNA copy are produced by the enzyme-catalysed reactions. But the relation between replicator and metabolism is more than one of parasitism. The DNA has evolved in such a way that it contributes to the maintenance of the metabolic cycle by selecting the adequate enzymes. That control function might also have been performed by a non-replicating selector, but replicating systems have a definite selective advantage compared to non-replicating systems (cf. Heylighen, 1992a,b), and so we may assume that possible non-replicating controls have been eliminated during the "struggle for life" amongst competing second-order metasystems. Some possible scenarios for the origin of life are discussed in this volume by Powers (this issue) and Umerez & Moreno (this issue).

Simple reflex

With the advent of living cells, characterized by a two level metasystem hierarchy, and a first level of control, as distinguished from inherent constraint, the time is ripe for a further development of control structures separate from the DNA itself. It is here that the rudiments of intelligence or mind appear. According to Turchin (1977), the first metasystem level above the level of life is *movement*, defined as the *control of position*. I have argued earlier that the appearance of the capacity to move is not a complete MST, but only a part of it (Heylighen, 1991c). Indeed, movement is merely a variation of the position, but there is as yet no constraint. It is only in Turchin's second stage, *irritability*, that a constraint appears: movement now becomes a function of the state of the environment as sensed by the organism.

This is the beginning of *perception*: the organism can discern a change in the environmental situation before that change has had the chance to affect the organism's metabolic cycle. It can then react to the implications of that change by moving to different surroundings if the implications are negative. A very primitive example of such a control system can be found in the bacterium described by Powers (1989 and this issue): it changes the direction of its movement randomly, but the pace of changes is fast when the situation is negative (absence of food, presence of poisons), and slow when it is positive.

This is a genuine third level metasystem: whereas the second order metasystem can only react to changes in its surroundings by changing its own cycle of processes, the third order system is capable to actively seek new surroundings, thus varying the

constraint imposed upon it by the environment. The first level metasystem (dissipative structure) is varied in a way constrained by DNA and environment, but this second level constraint is itself varied on the third level by moving to a different environment.

On the level of irritability, a change in the environment leads to a perception, which through a shorter or longer metabolic pathway leads to an adapted movement, or, more generally, *action* which changes the surroundings. This direct perception-action connection may be called the *simple reflex*, and the pathway it uses will be materialized in the form of nerves connecting sensors (sense organs) to effectors (muscles). Although in biology the boundary between plants and animals is much more fuzzy than one would naively expect, one way to capture the intuitive sense of an “animal” is as an organism capable of simple reflexes. (ignoring certain capabilities of motion in plants, such as phototropism).

The principle of selective variety tells us that more and more pathways will tend to emerge, in order to cope with a maximum variety of environmental perturbations. We may expect that at a certain moment different reflexes will act simultaneously, interfering with, or possibly even opposing, each other’s effects. Again the time is ripe for a higher order constraint, co-ordinating and steering the reflexes of the level below.

Complex reflex

Turchin (1977) called the next level of metasystem control the *complex reflex*. Variation of simple reflexes can be achieved by interconnecting the different pathways, so that a network or nervous system is formed. A stimulus received by one of the sensors will now not have a unique path to follow in order to stimulate a muscle to action: several possible pathways will be available, depending on the internal state of the nervous system, which is determined by the different stimuli received at the present and past moments. For example, a decrease of the external temperature sensed by the organism would previously have immediately triggered an action seeking a higher temperature (e.g. by moving up in the water). On the level of the complex reflex this perception will be first compared with possible other relevant informations. If, for example, the system is aware of a higher level of internal activity producing heat, or anticipates an external increase of temperature on the basis of different perceptions, it may decide not to start the warmth-seeking action.

The level of complex reflexes is also that of *variablegoals*. In the simple reflex, the goal or reference state that the system tries to achieve or maintain by reacting is implicit and fixed, e.g. keeping the system at a fixed external temperature. In the complex reflex, the goal becomes an explicit state of the nervous system that depends on the total sensory input. For example, the ideal temperature will be varied according to the physiological processes the organism is undergoing (e.g. higher reference temperatures during infections). This sensory input-dependent variation of goal or reference states is well described by Powers (1973; 1989). But the level of complex reflexes is a metasystem not

only with respect to the variation of specific goals. The complete pathway that an impulse follows through a nervous network will vary in a way constrained by the total input, internal state and global organization of the nervous system.

That constraint, however, is still fixed by the static structure of the nervous network. The same input combined with the same internal state will activate the same pathways, and produce the same actions, each time again. Although the state of activation of the network can be conceived as a short-term memory of previous perceptions, there is no long-term memory that would store newly learned or discovered patterns of behavior.

Associating

The next MST will create the capacity for learning, by making the connections between different possible pathways variable. This means that the likeliness that a signal would follow one path rather than another one will change under the effect of experience. The benefit of this variability is that it allows the organism to adapt its reactions to an environment that either is changing, or is so complex that adapting to it by a genetic alteration of the "hard-wired" nervous systems would take too much time. Again we see that the more variety the environment proposes, the higher the metasystem level that is best adapted to it.

The constraint governing the learning of new pathways depends on the reaction of the environment to the actions of the system. This reaction may either reinforce pathways leading to actions whose result is evaluated as adequate by the system, or weaken pathways leading to actions with results perceived as inadequate. There are also internal constraints, such as the Hebb rule which states that neurons that are activated simultaneously will develop stronger connections. The exact rules that constrain the change of neural network connections under experience are still being studied through methods such as connectionist computer simulations, psychological experiments, and brain research.

Remark that there is as yet no evidence of a separate control system that would monitor the behavior of connections, and initiate actions to change them when they are evaluated as inadequate. The constraint governing changes in associations rather seems to emerge from the dynamics of the global neuronal activation system in interaction with the environment. That is probably the reason why Powers (1973; 1989) does not include learning in his hierarchy of control levels, but rather views it an independent process.

Though we do not know the precise details about learning rules as yet, a general constraint appears to be that connections are formed between impressions that are contiguous or close in (mental) space or time. If stimulus A is followed closely by stimulus B, and this sequence is repeated, a neural connection will be formed between the patterns representing A and B, so that B will tend to become activated as soon as A is perceived. Normally, no connections will be formed between impressions that are widely separated. This is an adequate adaptation, since the connections are basically there to

anticipate causally linked events, and repeated temporal contiguity seems like the simplest indication of causal linkage. Yet, it still forms a restriction, and the system would be more adaptive if that constraint could be varied. But that requires another MST...

Rationality or thinking

The constraint that steers learning at the level of associating is based on contiguity: new associations will only develop if the patterns they connect are experienced together. This experience is determined by the environment: non-connected patterns are only jointly activated by joint perceptions of external phenomena. For example, repeatedly experiencing the visual phenomenon “flower” followed by the olfactory phenomenon “perfume”, will establish an association between “flower” and “perfume”, so that the cognitive system would expect or predict a perfume the moment a flower is perceived. On the other hand, on the level of learning no connection between “flower” and “music” would be established if these phenomena are never perceived together.

Yet, it is possible for us, human beings, to associate the two concepts, and to conceive different situations where flowers would be present together with music (e.g. Sixties rock concerts, a movie version of Beethoven’s “Pastorale”, or the musical background of a flower exhibition). Turchin (1977) has called this process “thinking”. It can be defined as the creation of associations in one’s mind between phenomena that may never have been associated in reality. It implies the variation of the constraint imposed by external contiguity upon learning. But a real MST demands a higher order constraint to steer this variation. On the level of thinking, the constraint is that the things that are freely associated belong to a restricted system of modular units obeying specific rules. I have called those units “concepts” (Heylighen, 1991c), and their sensory equivalents “symbols”.

The role of symbols is that of a concrete, sensory support for the abstract symbols. As our nervous system is not built for direct association between perceptually separate patterns, the symbols function as a kind of short-cutting connection: each abstract concept has a learned association with a perceptual symbol (e.g. a written or spoken word, an image, a sound, ...). The symbol can be perceptually associated with another symbol (e.g. by reading or hearing both symbols in contiguity), which is itself associated with an abstract concept, thus forming an indirect association between the two concepts.

The symbols and concepts belong to a learned lexicon or vocabulary, and their associations are governed by a number of learned grammatical, syntactical or logical rules (e.g. an “adjective” concept can be associated with a “noun” concept but not with a “conjunction” concept). The sets of units and rules developed basically through *cultural* evolution: patterns that turned out to be useful in combinations were transmitted and retransmitted between individuals and thus were retained in collective memory. Such replicating cognitive patterns are called “memes” (Dawkins, 1976; Moritz, 1990; this issue; Heylighen, 1992b). The important point is that memes will only be retained if they

satisfy a number of selection criteria (Heylighen, 1993). In addition to general criteria based on usefulness to the organism, a fundamental requirement is that they have a maximal invariance over contexts, i.e. the meaning of a symbol should vary minimally when used in different situations. It is this invariance of individual units (and rules) which makes a constrained variation of their combinations and associations possible. But the invariance of a particular pattern is only ascertained after many rounds of collective trial-and-error and transmission. Thus the development of new concepts and rules is a slow, uncontrolled process, that might be made much more efficient by yet another MST.

Metarationality

Turchin sees the metasystem controlling the level of thinking as simply culture, and the corresponding MST as the deepening and widening of that socio-cultural system into a real “superbeing”. Indeed, as we saw, it is cultural evolution that develops new units and rules of thought. But cultural evolution is a part of general evolution, and in the limit every system in the world is being varied by evolution. Thus we might call evolution the *ultimate metasystem*, controlling all other systems. (Turchin is prepared to equate such a “highest level of control” with the metaphysical concept of "God" (Turchin & Joslyn, 1990), although I would prefer the more neutral labels of “Evolution” or “Nature”.) This system is different from other metasystems, though, in the sense that it is impossible to determine any absolute and non-tautological constraint governing it: everything can be changed by evolution (or God, if you prefer), but no type of evolutionary change can be a priori excluded.

When trying to determine the nature of the MST following rationality, we therefore should ask whether culture can be characterized in some way by an invariant constraint. At present, I don’t see any restriction of this type. Apart from general principles of evolution, the only invariant rule governing culture seems to be that everything changes: no idea, theory or law is sacred. Widening and deepening the socio-cultural “being” by faster and more direct communication of memes between individuals would seem to only accelerate this unconstrained variation. My criticism of Turchin’s analysis is similar to the one I made earlier about his analysis of movement. In my definition, a metasystem implies a higher level variation together with a constraint on that variation. Both movement and cultural evolution can be viewed as variations of lower level constraints, but they both lack a higher level constraint.

Another possibility to determine a cultural MST would be to see culture as a separate system that *controls*, and not merely *constrains*, the systems at the level below. Again, I cannot see any physical separation (implying non-trivial communication channels) between people’s thoughts, and the culture that is supposed to control them. The constraints implied by Turchin’s view of social integration between individuals are merely additional constraints on the exchange of thoughts, not on the development of new systems of thinking. Thus, Turchin’s “superbeing” would in my definition be merely a

"supersystem" (albeit with enhanced variation) but not a "metasystem": constraints and variations are added in parallel, not in a hierarchical relation.

In order to make a true MST from the level of rationality we need systems (constraints) that vary the constraints governing thought, that is we need systems that efficiently develop new concepts, rules and models, which are more than mere associations of existing ones. This requires on the one hand a maximal variety of existing cognitive material that can be analysed, recombined and integrated in order to produce potential new models. The most obvious way to achieve this seems through the electronic revolution, which offers to put the whole of human knowledge, plus information gathered continuously by all possible sensing devices (satellites, robots, sensors, ...) at anyone's fingertips, via a world-wide network of interconnected computers. Fast computers can then generate all imaginable combinations and variations of this information as raw material for potential new models. On the other hand, the enormous complexity of this information will require smart systems for searching, selecting, structuring and reorganizing knowledge. These systems can be based on fundamental insights in the function and development of knowledge. Existing examples of such "metacognitive" algorithms are: neural networks, induction-based machine learning, genetic algorithms, statistical techniques for clustering and factor analysis.

A future integration (which is already under the way) of these different approaches, perhaps based on an overall theory of knowledge evolution, promises a mastery of model-building or discovery which is beyond anybody's imagination. A concrete effect will be that reliable models of complex systems (say, the solar system, a society, an ecology), which now demand multiple years of work by groups of the most smart scientists, will be generated for any individual in minutes or seconds. These models, implemented, executed and kept up-to-date by computers, will allow people to predict, manage and redirect systems which seemed for ever beyond control.

The question remains whether the new metasystem will have a scale larger than one or not. In other words, do we need social integration of different individuals into a collective metasystem ("superbeing"), or can a single individual become a metasystem ("metabeing")? At present, we do not seem ready to answer this question. In a separate paper (Heylighen & Campbell, this issue) I will review some arguments for and against the possibility or necessity of social integration. At present, I just want to note that none of the cognitive MST's we have reviewed (starting with the origin of life) seem to include any form of integration between initially separate systems (perhaps with the exception of the complex reflex). But, as it will be shown now, that is not the general case.

Some parallel branches

Though we have discussed MST's until now as if they form a single sequence, the possibility of parallel, bifurcating MST's was mentioned earlier. The principle can be easily demonstrated at the level of the origin of life: the (second level) constraint on the

variation of an organism's physiology has two components, DNA and environment, which can be varied independently of each other. Variation of the latter is what we discussed as movement. Variation of the former can be found in different mechanisms of genetic change: mutations, recombinations, copying errors, ... In order to find an MST, we need to establish a mechanism that constrains such changes in a non-random way. Although geneticists are still learning a lot about the underlying molecular processes, there is one mechanism which is clearly not the effect of random noise, *sexuality* It can be defined as the constrained variation of DNA sequences by "cross-over" with a second sequence from another organism of the same type. It makes it possible to increase genetic variety without dangerous side-effects. The metasystem that emerges from this MST is the *species*, which is defined as the set of all organisms that are capable to recombine their genes in the way mentioned. This parallel MST branch is characterized by a much slower, and less controlled pace of 3rd order variation than the one of movement, but has obviously played an essential role in evolution.

A second way to vary the constraint exerted by DNA, is to vary the activation of the different genes residing on the DNA string, so that different copies of the same DNA molecule could implement different "genetic programs" (Kauffmann, 1984; 1993). This requires a parallel development of cell types, each characterized by a different activation of a shared DNA template. This can be found in *multicellularorganisms*, where during embryological development the cells get differentiated through a differential activation of their genes. Note that this MST requires also a supersystem transition, integrating different cells in a single organism. Thus the ontogenetical development parallels (not necessarily recapitulates) the constrained variation of DNA in the phylogeneticalevolution sketched above. The ontogenetical variation is much faster than the phylogenetic one, yet slower than the variation controlled by the simple reflex, which is why we chose the latter to represent the "main sequence" of MST's.

Once organisms have reached one of the "mental" metasystem levels (reflexes, associating, rationality, ...), characterized by perception and action, they are capable to interact in a manner less direct than the one in which chemical signals are exchanged between contiguous cells. They can now can use different "actions at a distance", such as perceptual signalling or linguistic communication to co-ordinate their actions. If the result of their joint actions is more than the sum of their actions as individuals this interaction can be interpreted as a *co-operation* (Heylighen, 1992c), and the group of individuals then becomes a *social system*. The social constraint imposed upon individuals' behavior by their co-operative arrangement can be interpreted as producing a supersystem. In order to also have a metasystem, the constraints characterizing each individual should be varied too, like in the differential activation of genes in different cells. Whether this really happens is debatable (see Heylighen & Campbell, this issue), but we might at least conceive that individual *dos* and *don'ts* are somehow set or changed by society. This may lead to a division of labor, where different individuals carry out specialized tasks, and thus to an overall increase of variety for the social system. This MST seems again parallel

to the main sequence, although Turchin, as noted in the previous section, would argue that the social and the mental branches would merge in the ongoing MST, producing a “superbeing”.

Summary

The present paper has proposed to redefine Turchin’s original concept of a metasystem, starting from the definition of a “system” as a constraint on variety. By distinguishing constraint, static variety, and dynamic variety, this allowed us to build a 2 × 2 × 2 classification in which two classes of metasystems are distinguished from supersystems and other related types of emergent phenomena. Metasystems are then constraints on the variation of constraints on variety. This made it possible to situate the parallel theories of Turchin, Powers, Campbell and Simon into an integrated framework describing the evolutionary emergence of hierarchical systems.

We discussed some evolutionary principles describing how and why this emergence takes place, explaining the benefits of MST’s through the principle of selective variety. We analysed the positive feedback between variety and constraint underlying the “branching growth of the penultimate” level, concluding that MST’s should be viewed as the phases of rapid change in a continuous evolutionary movement toward increasing variety.

We then reviewed the most important MST’s in the history of evolution, while reinterpreting the resulting systems as constrained variations of lower level systems. This history can be summarized in the following scheme (similar to the scheme used by Turchin (1977), albeit with some differences):

mechanical motion	=	*constrained variation of*	physical state
dissipative structure	=	*constrained variation of*	mechanical motions
living cell	=	*constrained variation of*	dissipative structure(s)

(constrained by DNA structure + DNA activation + environmental situation)

multicellular differentiation	=	*constrained variation of*	DNA activation
sex	=	*constrained variation of*	DNA structure
simple reflex	=	*constrained variation of*	environmental situation
complex reflex	=	*constrained variation of*	simple reflexes
associating	=	*constrained variation of*	complex reflexes
thinking	=	*constrained variation of*	associating
metarationality	=	*constrained variation of*	thinking
social interaction	=	*constrained variation of*	behavior

(reflexes, associating, thinking)

Acknowledgments:

I wish to thank Valentin Turchin and Cliff Joslyn for many deep-going discussions about these and related subjects, shaping my thoughts on the issue. During this research I have been supported by the Belgian "National Fund for Scientific Research" (NFWO), as a "Senior Research Associate".

References

Campbell D.T. (1974): "Evolutionary Epistemology", in: The Philosophy of Karl Popper, Schilpp P.A. (ed.), (Open Court Publish., La Salle, Ill.), p. 413-463.

Conant R.C. & Ashby W.R (1970): "Every Good Regulator of a System Must Be Model of that System", Int. J. Systems Science 1:2, p. 89-97.

Dawkins R. (1976): *The Selfish Gene*, (Oxford University Press, New York).

Glueck R. & A. Klimov (this issue): "Applications and potential use of Metasystem Transition in Computer Science", World Futures.

Heylighen F. & D.T. Campbell (this issue): "Selection of Organization at the Social Level" (World Futures).

Heylighen F. (1989): "Self-Organization, Emergence and the Architecture of Complexity", in: *Proceedings of the 1st European Conference on System Science*, (AFCET, Paris), p. 23-32.

Heylighen F. (1990): "Relational Closure: a mathematical concept for distinction-making and complexity analysis", in: *Cybernetics and Systems '90*, R. Trappl (ed.), (World Science, Singapore), p. 335-342.

Heylighen F. (1991a): "Structuring Knowledge in a Network of Concepts", in : *Workbook of the 1st Principia Cybernetica Workshop,* Heylighen F. (ed.) (Principia Cybernetica, Brussels-New York), p. 52-58.

Heylighen F. (1991b): "Modelling Emergence", *World Futures: the Journal of General Evolution* 31 (Special Issue on *Emergence*, edited by G. Kampis), p. 89-104.

Heylighen F. (1991c): "Cognitive Levels of Evolution: pre-rational to meta-rational", in: *The Cybernetics of Complex Systems - Self-organization, Evolution and Social Change,* F. Geyer (ed.), (Intersystems, Salinas, California), p. 75-91.

Heylighen F. (1992a): "Principles of Systems and Cybernetics: an evolutionary perspective", in: *Cybernetics and Systems '92*, R. Trappl (ed.), (World Science, Singapore), p. 3-10.

Heylighen F. (1992b) : "Evolution, Selfishness and Cooperation", *Journal of Ideas,* Vol 2, # 4, pp 70-76.

Heylighen F. (1992c) : "'Selfish' Memes and the Evolution of Cooperation", *Journal of Ideas* , Vol. 2, #4, pp 77-84.

Heylighen F. (1993): "Selection Criteria for the Evolution of Knowledge", in: *Proc. 13th Int. Congress on Cybernetics* (Association Internat. de Cybernétique, Namur), p. 524-528.

Joslyn C. (this issue): "Semantic Control Systems", World Futures.

Kauffman S.A. (1984): "Emergent Properties in Random Complex Automata", Physica D 10, p. 145.

Kauffman S.A. (1993): Origins of Order: self-organization and selection in evolution, (Oxford University Press, Oxford).

Mesarovic M.D. & Takahara Y. (1975): General Systems Theory : Mathematical Foundations, (Academic Press, New York).

Mesarovic M.D., Macko D. & Takahara Y. (1970): Theory of Hierarchical Multilevel Systems, (Academic Press, New York).

Moritz E. (1990): "Memetic Science: I - General Introduction", Journal of Ideas 1, p. 1-23.

Moritz E. (this issue): "Metasystem Transitions, Memes, and Cybernetic Immortality", World Futures (this issue).

Nicolis, G, and Prigogine, Ilya (1977): S*elf-Organization in Non-Equilibrium Systems,* (Wiley, New York).

Pattee, Howard: ed (1973): Hierarchy Theory, (George Braziller, New York).

Powers, W.T. (1973): B*ehavior: the Control of Perception,* (Aldine, Chicago).

Powers, W.T. (1989): "An Outline of Control Theory", in: *Living control systems,* Powers W.T. (ed.), (Control Systems Group, Gravel Switch: KY), p. 253-293.

Powers W.T. (this issue): "The Origins of Purpose: the first Metasystem Transitions" World Futures .

Simon H.A. (1962): "The Architecture of Complexity", *Proceedings of the American Philosophical Society* 106, p. 467-482.

Turchin V. (this issue): "A Dialogue on Metasystem Transition", World Futures.

Turchin V. and Joslyn C. (1990): "The Cybernetic Manifesto", *Kybernetes* 19, Nos. 2-3, p. 63-65.

Turchin, Valentin (1977): *The Phenomenon of Science. A Cybernetic Approach to Human Evolution,* (Columbia University Press, New York).

Umerez J. & A. Moreno (this issue): "Origin of Life as the First Metasystem Transition. Control Hierarchies and Interlevel Relation", World Futures.